MISKOLCI EGYETEM DOKTORI (PH.D.) TÉZISFÜZETEI
HATVANY JÓZSEF INFORMATIKAI TUDOMÁNYOK DOKTORI ISKOLA

A megerosítéses tanulás és a szimulált hutés kombinált használata: algoritmusok és alkalmazások

Készítette:

STEFÁN PÉTER

okleveles mérnök-informatikus

okleveles közgazdász

AKI DOKTORI (PH.D.) FOKOZAT ELNYERÉSÉRE PÁLYÁZIK

Miskolc-Budapest, 2003

A Bíráló Bizottság tagjai

Elnök:

Dr. Galántai Aurél	a matematikai tudomány kandidátusa

Titkár:

Dr. Szalay Tibor	Ph.D.

Tagok:

Dr. Erdélyi Ferenc	a muszaki tudomány kandidátusa
Dr. Radeleczki Sándor	a matematikai tudomány kandidátusa
Dr. Rudas Imre	a muszaki tudomány kandidátusa

Hivatalos bírálók:

Dr. Kovács Szilveszter	Ph.D.
Dr. Váncza József	a muszaki tudomány kandidátusa

1. Bevezetés

A mesterséges intelligencia tudományágának napjainkban egyik legdinamikusabban fejlodo területe az elosztott, ágens-alapú rendszerek problématerülete. E rendszerek legfontosabb célkituzése, hogy olyan modelleket építsenek föl, amelyben egy adott probléma megoldását egymással elsosorban mellérendeltségi viszonyban álló, intelligens ágensek, egymással kommunikálva, együttmuködve végzik. E rendszereket, mivel vagy nincsenek benne hierarchikus viszonyok, vagy azok dinamikusan jönnek létre és bomlanak le, heterarchikus rendszereknek is nevezik.

Az ágens-alapú rendszerek felépítése során két alapveto feladatot kell vizsgálni:

1. az ágensek belso felépítését, valamint
2. az ágensek közötti kommunikációs tevékenységet.

Az ágens, belso felépítését tekintve, legfontosabb tulajdonsága az autonómia, amely azt jelenti, hogy az ágens önállóan, korábban kívülrol származó, de alapvetoen leképzett, belso információi alapján, önálló döntéseket hoz. E tulajdonság sugallja azt, hogy az ágens a rendszerben egy „cselekvo” jellegu elem, akciókat hajt végre, amelyeket közvetve vagy közvetlenül visszahatnak az ágensre, például jutalom vagy büntetés formában. Az ágens feladata az, hogy ezeket az információkat figyelembe vegye, és felhasználja annak érdekében, hogy az adott feladatot, egyre pontosabban, a környezet elvárásaihoz leginkább alkalmazkodva hajtsa végre, azaz tanuljon.

A megerosítéses tanulás egy olyan gépi tanulási módszer, mely leginkább alkalmazkodik egy döntéshozó ágens autonómiájához. Szemben a passzív tanulási módszerekkel, ahol egy külso cselekvo „instruálja” a tanuló rendszert, a megerosítéses tanulás aktív módszer, ami azt jelenti, hogy az ágens kezdeményezi a tanulást, adott szituációban cselekszik, érzékeli (méri) a környezetben a cselekvés hatásait, és ezeket a hatásokat kiértékeli, a saját „tudásaként” szintetizálja.

Az ágens legfontosabb jellemzoje a döntéshozatal, melynek idobeli szekvenciáját az ágens stratégiájának („*policy*”) nevezik. Alapvetoen kétféle döntési stratégia létezik: felderítés („*exploration*”), azaz új lehetséges megoldások feltérképezése, valamint a kiaknázás („*exploitation*”), azaz a már megismert megoldások felhasználása. A stratégia lehet idoben

állandó, vagy idoben változó. A megerosítéses tanulás egyik nagyon fontos kérdése: hogyan lehet e két szélsoséges stratégiát idoben úgy megváltoztatni, hogy a változó stratégia az ágens, környezetbol szerzett várható honoráriumának összegét maximalizálja.

A megerosítéses tanulási módszerek alapgondolata a dinamikus programozásban gyökerezik. E tématerület legkiválóbb alakja *Richard Bellmann*, aki elvi alapokon igazolta, hogy egy formalizált problémára, a környezeti dinamika ismeretében lehet idoben elhatárolt döntésekkel optimális megoldást találni. A megerosítéses tanulás alapvetoen dinamikus programozási feladatot old meg, de a környezet dinamikájának pontos ismerete nélkül, mintavételezéssel és becsléssel. E területen *Andrew Barto* és *Richard Sutton* ért el kiemelkedo eredményeket, kidolgozva az idobeli-differenciák módszerét. E tanulási módszer gyakorlati alkalmazásában többek közt *Leslie Kaelbling* valamint *Michael Littman* ért el jelentos sikereket.

A megerosítéses tanulás, valamint a heterarchikus rendszerek vizsgálata Magyarországon az elmúlt években lett „divatos” kutatási irányzat. A tématerület legfobb hazai kutatói *Lorincz András*, valamint *Szepesvári Csaba*, akik számos matematikai tétel kidolgozásával gazdagították a megerosítéses tanulási módszereket.

A mesterséges intelligencia üzleti és gyártási folyamatokban való alkalmazásának legfobb hazai úttöroi *Márkus András*, *Monostori László*, *Váncza József*, *Kárár Botond* valamint *Viharos Zsolt János*. Az o nevükhöz fuzodik számos kombinált mesterséges intelligencia alkalmazás (neuro-fuzzy rendszerek, input-output kereso rendszerek, genetikus algoritmuson alapuló ütemezok) kialakítása, valamint a heterarchikus elven muködo gyártórendszerek modelljének továbbfejlesztése.

2. A dolgozat célkituzése, a kutatás módszere

A dolgozat célkituzése többszörös: egyrészt a döntéshozó ágens számára olyan felderítés-kiaknázás egyensúlyozási módszer kidolgozása, mely elméletileg kelloképp megalapozott, és a gyakorlatban is jól használható, másrészt olyan algoritmusok készítése, melyek heterarchikus modellt használnak olyan problémák megoldására, mint az Internet útvonal-irányítási feladat (Internet Protocol Packet Routing Task), vagy a futószalag ütemezési probléma (Flow-shop Scheduling Problem).

A kutatási munkám az alábbi körfolyamattal jellemezheto:

1. az adott résztémához kapcsolódó anyagok összegyujtése, megismerése, szintetizálása,
2. a feladathoz kapcsolódó probléma megfogalmazása,
3. a problémára matematikai modell, algoritmus, illetve heurisztika kidolgozása,
4. számítógépes implementáció, a kidolgozott algoritmus összehasonlítása más, klasszikus megoldásokkal,
5. tesztelés, kiértékelés,
6. majd az egész folyamat részleges ismétlése a probléma-feltárás részletezettebb szintjén.

Kutatómunkámat a *Miskolci Egyetem Alkalmazott Informatika Tanszékén*, valamint a *Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutatóintézetének* (MTA SZTAKI) *Intelligens Gyártási és Üzleti Folyamatok Laboratóriumában* (IMBP) *végeztem*.

3. A feladat megoldásának módszerei

A feladat megoldása során az alábbi egyszeru ágens-modellt használtam: adott szituációban az ágens számára rendelkezésre áll egy akció-halmaz, amibol a döntés folyamán kiválaszt egyet. Az akcióhalmaz minden eleméhez hozzárendelünk egy valószínuség-értéket, ami azt jelzi, hogy mennyire valószínu az adott elem kiválasztása. Ez az alábbi, ún. Boltzmann-képlet segítségével teheto meg:

$$p(a_i) = \frac{e^{\frac{Q_i}{T}}}{\sum_{\forall a_j \in A} e^{\frac{Q_j}{T}}}.$$

Itt a $p(a_i)$ az i. akció választásának valószínusége, A az akcióhalmaz, Q_i az i. akció preferenciája, és T a szimulált homérséklet, mint szabályozó paraméter.

E paraméter megfelelo megválasztásával a két korábban említett karakterisztikus ágens-stratégia a felderítés és a kiaknázás elérheto: ha a T kelloképp magas (tart a végtelenhez), a definiált eloszlás egyenletes, azaz minden akció választásának valószínusége egyforma. Ez nem más, mint a felderítés. Ha viszont a T kelloképp kicsi (tart a nullához), akkor a definiált eloszlás, ún. *„mohó eloszlás”*, azaz a legpreferáltabb akció választása biztos esemény, míg a többi akció választásának valószínusége zéró. (Ha a preferenciák diszkrét értékek,

elofordulhat, hogy több egyforma, maximális értékünk van. Ez esetben a biztos esemény „*1-valószínusége*” egyenlo arányban megoszlik a legjobb lehetoségek között.) A viselkedés megfelel a kiaknázás stratégiának, azaz biztosan a legnagyobb preferenciával rendelkezo akciót választja az ágens.

A Boltzmann-képlet e tulajdonságai matematikai formába önthetok és bizonyíthatók. A kérdés az, hogy a T változtatásával mennyire közelítjük meg a két szélsoséges stratégiát? Pontosabban a kérdés fordítva teheto fel: léteznek-e olyan véges és nem-nulla homérséklet értékek, mely mellett a definiált eloszlás egy bizonyos hibahatáron belül egyenletes, illetve „*kelloképp*”, egy bizonyos hibahatáron belül, „*mohó*”. Munkám során elméleti levezetéssel igazoltam, hogy amennyiben a preferencia-értékek végesek, úgy ezek a határok léteznek, végesek, és egyszeru matematikai módszerekkel meghatározhatók.

A véges homérséklethatárok birtokában, valamint ha az ágens tudja, hogy mennyi idot szeretne eltölteni felderítéssel, definiálható egy hulési ütemterv egyszeru kétpontos interpolációval úgy, hogy az ágens stratégiája a kezdeti idopillanatban felderítés (a hozzá tartozó homérséklettel), és kiaknázás a (felfedezésre szánt) végso idopontban. E két szélso eset között folyamatos átmenet húzható, ily módon az ágens próbál annyi információt összegyujteni kezdetben, amennyi lehetséges, majd fokozatosan „*beszukíti látókörét*” a legjobb addig megtalált döntés, vagy döntések köré.

E módszer használható hálózati útvonal-irányítási feladatok esetén, amikor minden útvonal-irányító eszköz („*router*”) egy-egy ágens, és arról hoz döntést, hogy egy beérkezo adatcsomag (adatcsomagok halmaza) melyik kimeno hálózati interfészén haladjon tovább, míg a célját el nem éri. Mivel a hálózat gyakran változik, ráadásul a változás a terheléstol is függ, célszeru az ágenseknek az elobbi felderítés-kiaknázás metódust használni a hálózat változó részeiben. Az általam kidolgozott útvonal-irányítási algoritmus kihasználja az irodalomban népszeru, de a gyakorlatban kevéssé használt Q-forgalomirányítási algoritmus összes adaptív elonyét, valamint a hulési modell segítségével alkalmassá teszi azt az adattovábbítási hurkok „*megbüntetésére*”, valamint egy régi jól muködo útvonal visszaállítására („*path recovery*”).

A kutatás kiterjedt nemcsak az ágens belso szerkezetének leírására, hanem az ágensek közötti strukturált információcsere (protokoll) kidolgozására is.

Az útvonal-irányítási probléma matematikai vetülete a legrövidebb-út probléma, ami polinomiális idoben futó algoritmusokkal megoldható. Hasonló, de minoségileg bonyolultabb probléma a futószalag-ütemezés („*flow-shop scheduling*”), mely tipikus példája az NP-teljes feladatoknak. A megerosítéses tanulás szempontjából azonban rendkívül hasonló a legrövidebb-útvonal keresési problémához, azzal a lényeges különbséggel, hogy a döntési

lehetoségek az ütemterv kialakítása során elfogynak. Emiatt a megerosítéses tanulás néhány matematikai elofeltétele sem teljesül. Mérésekkel igazoltam, hogy a megerosítéses tanulás még ilyen *„mostoha körülmények”* között is muködoképes, megfelelo heurisztikai elemek bevonásával, és képes a klasszikus heurisztikus megoldásoknál kisebb veszteségi-ideju ütemtervet generálni. Mindez egy valós gyártórendszer-virtuális gyártórendszer keretrendszerben valósul meg, mely alkalmassá teszi az algoritmust a környezetben lezajló változások érzékelésére, és a változásoknak megfelelo új feladat-sorrend kialakítására. Munkámban rávilágítok arra is, hogy a megelozési-relációkat tartalmazó eset gyengíti az ütemezési problémát, azaz kevesebb számítással adható megfelelo eredmény. A *„megfelelo”* szó jelen esetben kvázi-optimális megoldásokra utal.

4. Az értekezés tézisei

Az ágens-alapú rendszerek valamint a megerosítéses tanulás területén végzett kutatásaim eredményei három fo pontban foglalhatók össze.

1) Kidolgoztam egy szimulált hulést használó eljárást, mely a választható akciók halmazának preferencia-értékei fölött, a Boltzmann-képlet alkalmazásával definiált valószínuség-eloszlás segítségével alkalmas a megerosítéses tanulást használó ágens felderítés, illetve kiaknázás stratégiái közötti egyensúlyozás megvalósítására. Ennek keretén belül:
   a) Igazoltam, hogy amennyiben a Boltzmann-képlet homérséklet paramétere tart a nullához, a definiált eloszlás tart a *„mohó”* eloszláshoz, amennyiben a szimulált homérséklet tart a végtelenhez, a definiált eloszlás tart az egyenletes eloszláshoz.
   b) Meghatároztam azokat a legnagyobb/legkisebb homérséklet értékeket, amelyen túl/amely alatt az egyenletes eloszlás/mohó eloszlás eloírt hibahatáron belül megközelítheto.
   c) A homérséklethatárok, valamint a felderítésre szánt idotartam segítségével, egyszeru interpolációs technikával hulési ütemtervet vezettem be.
   d) Az eredményeket számítógépen ellenoriztem az irodalomban elterjedt *„n-karú rabló”* probléma segítségével.

2) Az elso tézis eredményeit felhasználva megerosítéses tanuláson alapuló, elosztott hálózati forgalomirányítási algoritmust készítettem, amely alkalmas útvonal újratanulásra, valamint az adattovábbítási láncban keletkezett ciklusok észlelésére és elkerülésére.
   a) Beláttam, hogy amennyiben a Q-tanulás alapfeltételei fennállnak, úgy az adatstruktúrájának elosztott jellege a tanulás konvergenciáján nem változtat.
   b) Kialakítottam egy elosztott algoritmust, mely alkalmas a csatolt élhalmazon keletkezett hurokrészek büntetésére, és kiiktatására, miközben a hurkot nem tartalmazó útvonal-részek értékelése változatlan.
   c) Kialakítottam a hulési modell *„inter-ágens”* protokollját, és rámutattam arra, hogy ez miként illesztheto bele egy már létezo hálózati szintu protokollokba, mint amilyen például az Internet Protocol version 4.
   d) Kidolgoztam egy elosztott útvonal-irányítás szimulátort, mely alkalmas az eredmények ellenorzésére.

3) A második tézisben kimunkált módszerhez hasonló eljárást alkalmaztam a futószalag ütemezési feladatok megoldására virtuális gyártórendszer környezetben. Mérésekkel igazoltam, hogy a módszer kisebb megmunkálási holtidoket tartalmazó ütemterveket készít, mint a klasszikus heurisztikus módszerek, illetve az új környezet felépítésének köszönhetoen képes a környezet változásainak érzékelésére, és a megváltozott körülményeknek a réginél jobban megfelelo, új ütemterv kialakítására.
   a) A Boltzmann-képlet felhasználásával feladat sorbarendezési algoritmust készítettem, amely figyelembe veszi a feladatok közötti lehetséges megelozési-relációkat.
   b) Kidolgoztam egy a futószalag ütemterv kiértékelésére alkalmas eljárást, mely tetszoleges számú megmunkáló gép és tetszoleges számú feladat esetén, megadja az ütemterv összes megmunkáló gépen kialakuló veszteségi idoinek összegét, azaz az ütemterv „jóságát”.
   c) Kidolgoztam egy az a) és b) pontokban említett eljárásokat felhasználó megerosítéses tanulási algoritmust, melyet számítógépen implementáltam, és a segítségével kapott eredményeket összehasonlítottam klasszikus heurisztikus eljárások eredményeivel.
   d) Az eredményeket virtuális gyártórendszer-valós gyártórendszer környezetbe ültettem, mely a virtuális és valós részek közötti felületen definiált protokoll segítségével lehetové teszi az idoben lezajló változások figyelembe vételét, új ütemterv kialakítását, a már létezo ütemterv finomítását.

5. Továbbfejlesztési elképzelések

A tézisekben összefoglaltak a gyakorlatban is jól hasznosíthatók, egy lépéssel közelebb visznek a nagyobb rendszerek ágens-alapokra helyezéséhez.

Elméleti síkon érdekes lehet, hogy a homérséklethatárok hogyan határozhatók meg, ha kilépünk a valós számok világából, és komplex számokkal dolgozunk.

A forgalom-irányítási algoritmus több irányban is továbbfejlesztheto: egyrészt megvizsgálható, hogy az Internet Protocol 4-es verziójánál (IPv4) korszerubb protokollok esetén, például az IPv6 esetén miként valósítható meg az algoritmus. Érdekes, a gyakorlathoz közel álló feladat az új protokoll, klasszikus, dinamikus útvonal-irányítási keretrendszernek megfelelo implementálása, külön e célra létrehozott, a normál forgalom-irányítási adatstruktúráktól elkülönített „*felderíto forgalom-irányító táblák*” segítségével.

A legperspektivikusabb ötletek az ütemezéshez kapcsolódnak. A jövobeli kutatási elképzeléseket az alábbi kérdések irányítják: Hogyan lehet más, nem futószalag ütemezési feladatot a kidolgozott keretrendszerben implementálni? Hogyan lehet minél hibaturobb, minél robusztusabb ütemezo-ágenst kialakítani? Hogyan alkalmazható az algoritmus meghatározott idohorizontra (feladat-horizontra) eloretekinto változata teljesen folyamatos üzemmódban?

6. Az értekezés témájában megjelent tudományos közlemények

Angol nyelven megjelent közlemények:

[P1] **Stefán, P.**; Monostori, L.: Shop-floor scheduling based on reinforcement learning algorithm, 3rd CIRP International Seminar on Intelligent Manufacturing, ICME 2002, Ischia, Italy, 2002, pp. 71-74.

[P2] **Stefán, P.**; Monostori, L.; Vaskó, Z: Quasi-optimal solution to the traveling salesman's problem in variable environment, 3rd international conference of Ph.D students, University of Miskolc, 2001, pp 415-422.

[P3] **Stefán, P.**; Monostori, L.: On the relationship between learning capability and the Boltzmann-formula, Engineering of Intelligent Systems, Lecture Notes in AI 2070, IEA/AIE-01, 14th International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems, Budapest, Hungary, June 4-7, 2001, Springer, pp. 227-236.

[P4] **Stefán, P.**; Monostori, L.; Erdélyi, F.: Reinforcement learning for solving shortest-path and dynamic scheduling problems, 3rd International Workshop on Emergent Synthesis, IWES'01, Bled, Slovenia, 2001, pp. 83-88.

[P5] **Stefán, P.**; Monostori, L.: On Internet routing problems in dynamically changing environment, MicroCAD International Meeting on Information Technology and Computer Science, University of Miskolc, Hungary, 2001, pp. 221-216.

[P6] Monostori, L.; Kádár, B.; Viharos, Zs.J.; **Stefán, P.**: AI and machine learning techniques combined with simulation for designing and controlling manufacturing processes and systems, Preprints of the IFAC Symposium on Manufacturing, Modeling, Management and Supervision, MIM 2000, Patras, Greece, 2000, pp. 167-172.

[P7] Monostori, L.; Kádár, B.; Viharos, Zs.J.; Mezgár, I.; **Stefán, P.**: Combined use of simulation and AI/machine learning techniques in designing manufacturing processes and systems, Proceedings of the 2000 International CIRP Design Seminar on Design with Manufacturing: Intelligent Design Concepts Methods and Algorithms, Haifa, Israel, 2000, pp. 199-204.

[P8] **Stefán, P.**; Monostori, L.; Pupp, Z.: Reinforcement learning methods in information engineering, MicroCAD International Meeting on Information Technology and Computer Science, University of Miskolc, Hungary, 2000, pp.

[P9] Pupp, Z.; **Stefán, P.**: Novel applications of self-organizing maps in information technology problems, MicroCAD International Meeting on Information Technology and Computer Science, University of Miskolc, Hungary, 2000, pp.

[P10] **Stefán, P.**; Monostori, L.; Erdélyi, F.: Using symbolic and sub-symbolic methods in solving problems difficult to analyze, MicroCAD International Meeting on Information Technology and Computer Science, University of Miskolc, Hungary, 1999, pp. 195-200.

Magyar nyelven megjelent közlemények:

[P11] **Stefán, P.**: Megerosíto tanulási módszerek alkalmazása Internet Routing problémák megoldására, Fiatal Muszakiak V. Tudományos Ülésszaka, Bolyai Egyetem, Kolozsvár, Románia, 2000, pp. 1-4.

[P12] **Stefán, P.**: Szimbolikus és szub-szimbolikus módszerek az analitikailag nehezen kezelheto problémák megoldásában, Fiatal Muszakiak IV. Tudományos Ülésszaka, Bolyai Egyetem, Kolozsvár, Románia, 1999, pp. 137-140.

[P13] **Stefán, P.**: Megerosíto tanulási módszerek alkalmazása az informatikában, Doktoranduszok fóruma, Miskolci Egyetem, 1999, pp. 65-70.